Misfolded protein propagation in an integrated computational model of structural network and *LRRK2* gene expression

Wuchao Yan[1], Chenfei Ye[2], Tong Wang[1], Junyan Sun[3,4], Tao Wu[3,4], and Ting Ma[1,2,4,5]*, *Member IEEE*

***Abstract*—It has been widely accepted that Parkinson's disease (PD) is triggered and shaped by propagation of misfolded α-synuclein. Converging neurophysiological evidence suggests that leucine-rich repeat kinase 2 (*LRRK2*) is involved in membrane transport of PD pathogenesis. This study proposed an agent-based computational model by integrating structural connections and gene expression to investigate whether *LRRK2* would affect the PD pathology propagation in central nervous system. Gene expression profiles from the Allen Human Brain Atlas (AHBA) and multimodal brain MRI images from Parkinson's Progression Markers Initiative (PPMI) and Human Connectome Project (HCP) were employed for the model construction. The model results exhibit the involvement of *LRRK2* gene expression remarkably elevated model fitting (r = 0.73) compared with the traditional susceptible-infected-removed (S-I-R) model (r=0.60). Specifically, our model revealed that *LRRK2* is more likely to modulate pathology secretion out of neurons, rather than spreading into neurons. The findings support the theory of *LRRK2* gene expression modulating cell-to-cell propagation of misfolded proteins. As a result, the proposed model would bring new insights of understanding PD mechanism in terms of misfolded α-synuclein propagation.**

I. Introduction

α-synuclein is a major protein component of Lewy bodies, which are characteristic structures of PD pathology. Evidence from molecular, animal and human postmortem studies [1,2] points to misfolded neurotoxic proteins that propagate through the central nervous system via neuronal connections. Misfolded proteins can deposit as insoluble aggregates and gradually spread through synaptic connections to interconnected groups of neurons [3]. These pathogenic misfolded disease-specific proteins have a key role in the pathogenesis of both familial and sporadic PD [4]. Therefore, the study of the transmission model of misfolded protein can reveal the pathological mechanism of PD.

Recently, an agent-based spreading Susceptible-Infected-Removed (S-I-R) model was proposed based on brain networks [3], which integrates structural connectivity, functional connectivity, and gene expression to reproduce the spatiotemporal pattern of empirical atrophy in Parkinson's disease. The model revealed a key role for both connectome topology and geometry in shaping the distribution of atrophy and demonstrated that the genes α-synuclein (*SNCA*) and the glucocerebrosidase (*GBA*) transcription affected the concentration of α-synuclein and the vulnerability of the local region. Such modeling work provides a good platform to investigate different influential factors in disease development individually. Furthermore, modified model by integrating other mechanisms would reveal more underlying mechanisms of PD development, such as impact from LRRK2 gene expression.

Converging lines of evidence suggest that the *LRRK2* is involved in the pathogenesis of PD [5,6,7]. The expression of *G2019S LRRK2* selectively attenuates the defense of neurons that have strong defenses against the α-synuclein pathology [6]. Even in idiopathic PD patients, *LRRK2* kinase activity is increased [7]. The *LRRK2* is mainly a cytoplasmic protein, particularly relating to membrane organelles, such as mitochondria, endoplasmic reticulum, Golgi apparatus, endosomes, and synaptic vesicles [8]. These organelles play a key role in activity of proteins in cells, which in turn affects the infection rate of misfolded proteins. All the evidences imply that LRRK2 is an impact factor of PD development, especially in activity of misfolded proteins. In order to further discover the mechanism of the *LRRK2* gene expression on α-synuclein transmission and PD pathology, this study proposed a computational model based on the previous agent-based spreading SIR model [3] integrated with *LRRK2*. The modified model is expected to provide more precise simulation on misfolded protein propagation in PD development.

II. Method

A. Data acquisition

Brain anatomical MRI data of 234 PD patients and 114 healthy controls were obtained from Parkinson's Progression Markers Initiative (PPMI) [9] to capture brain atrophy in protein propagation modeling. Additionally, diffusion MRI data from 203 healthy participants in Human Connectome Project (HCP) [10] and gene expression profiles from the Allen Human Brain Atlas (AHBA) [11] were also extracted to model the brain structural network and genetic effects.

Research was supported by grants from National Key Research and Development Program of China (2018YFC1312000), the National Natural Science Foundation of China (61801145), and the Basic Research Project of Shenzhen Science and Technology Program (JCYJ20180306171800589)..

1. Department of Electronic and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, Guangdong, China

2. Peng Cheng Laboratory, Shenzhen, Guangdong, China

3. Department of Neurobiology, Neurology and Geriatrics, Xuanwu Hospital of Capital Medical University, Beijing Institute of Geriatrics, Beijing, China

4. National Clinical Research Center for Geriatric Disorders, Xuanwu Hospital Capital Medical University, Beijing, China

5. Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing, China.

(Corresponding author: Ting Ma. Phone: +86-755-26033608; fax: +86-755-26033608; e-mail: tmahit@outlook.com)

B. Image pre-processing

Brain parcellation: the brain regions used in this study included 68 cortical [12] and subcortical brain regions from Desikan-Killiany atlas [13], as well as the substantia nigra in the ATAG atlas [14]. Only the left hemisphere postmortem brain data in the AHBA was used to simulate transmission of neuropathology.

Structural connectivity: diffusion MRI data from HCP were used to construct anatomic network of α-synuclein. We use probabilistic white matter fiber tracking to rebuild adjacency matrix. The fiber number matrix was considered as the connection strength matrix of 42 brain area and the length of each brain interval matrix replaced a communication path length. Selecting the most often appear side (the probability of at least this edge appear being greater than 98%) and the fiber length is more than 10, we construct a group of level structure connection matrix.

Gene expression: mRNA transcription (measured using in-situ hybridization) profiles of *SNCA, GBA,* and *LRRK2* were averaged across samples in the same brain parcel and across the 6 subjects in the AHBA dataset. *SNCA* and *GBA* gene expression profiles determine the local synthesis and degradation of α-synuclein, while *LRRK2* gene expression profiles mainly regulate the transmission probability of α-synuclein.

Empirical atrophy: empirical atrophy is obtained from individual T1-weighted MRI scans of PPMI cohort. The deformation-based morphometry (DBM) in the brain regions of each participant (patient or healthy control group) was used to quantify the local volume change. Unpaired t-test was further performed between the patient and healthy control group, with 5000 multiple comparison corrections. The resulting t-statistic was used as a measure of the region's empirical atrophy [15]. Furthermore, the independent component analysis (ICA) of DBM was conducted, whereas the component number was set as 30 to obtain the PD-ICA network.

C. Model construction

In the SIR model proposed in this paper, the agents are individual proteins. These proteins are split into "S," the portion yet to be infected (normal proteins), "I," the portion capable of transmitting the infection (misfolded proteins), and "R," the portion no longer active in the spreading (metabolized and cleared proteins) [3].

SNCA gene expression from AHBA was mapped to the protein production rate per unit volume per unit time α_i (the synthesis rate in region i) by sigmoid function, and *GBA* gene expression is mapped to the protein clearance rate per unit time β_i(the clearance rate in region i).

$$\Delta N_i = \alpha_i V_i \Delta t - (1 - e^{-\beta_i \Delta t}) N_i \tag{1}$$

Formula (1) represents the increment of normal protein during Δt time (N_i represents the number of normal proteins in region i, V_i represents the volume of region i). After the normal proteins in each brain region have stabilized (with an error tolerance of less than 10^{-7}), misfolded proteins are added to infect each brain region with the SIR model. Formula (2) and (3) represents the increment of the two proteins during Δt time (M_i represents the number of misfolded proteins in region i).

$$\Delta N_i = \alpha_i V_i \Delta t - \left(1 - e^{-\beta_i \Delta t}\right) N_i - \left(e^{-\beta_i \Delta t}\right)\left(1 - e^{-r_i^0 M_i \Delta t}\right) N_i \tag{2}$$

$$\Delta M_i = \left(e^{-\beta_i \Delta t}\right)\left(1 - e^{-r_i^0 M_i \Delta t}\right) N_i - (1 - e^{-\beta_i \Delta t}) M_i \tag{3}$$

As in formula (4), r_i^0 is the baseline propagation rate, where L_i is the mapping of *LRRK2* gene expression into probability after normalization, representing the influence factor of *LRRK2* gene.

$$r_i^0 = \frac{1}{V_i L_i} \tag{4}$$

Agents in region i can stay in region i or enter the edge according to the multinomial distribution of probability ($\frac{w_{ij}}{\sum_j w_{ij}}$) per unit time , where $w_{ij}$is the connection strength of edge (i, j) (fiber tracts density between region i and j).

The *LRRK2* regulate the probability of staying region i or entering region i from the edge to indirectly reveal the membrane transport process of proteins. Formula (5) to (8) model the propagation of normal and misfolded α-synuclein. θ_i is the intrinsic influence parameter of proteins entering and leaving the brain region i,l_{ij} is the length of the connection between brain region i and brain region j.

$$P_{ri-ri} = \theta_i + (1 - \theta_i) L_i \tag{5}$$

$$P_{ri-e(i,j)} = (1 - P_{ri-ri}) \frac{w_{ij}}{\sum_j w_{ij}} \tag{6}$$

$$P_{e(i,j)-e(i,j)} = 1 - \frac{1}{l_{ij}} \tag{7}$$

$$P_{e(i,j)-rj} = \frac{1}{l_{ij}} \tag{8}$$

The neuronal death was modeled as the result of two processes: the accumulation of naturally misfolded proteins and the direct toxicity of afferent deletions (reduced neural signal input) of neuronal death in adjacent (connected) regions, as in formula (9), where $\Delta A_i(t)$ is the amount of atrophy in brain region i at time t and $r_i(t)$is the proportion of misfolded agents in region i at time t . k_1 and k_2 are coefficients that measure the two processes, and their sum is 1. The main cause of cognitive dysfunction with multisystem atrophy (MSA) is the loss of the prefrontal striated afferent nerve, and the pathogenesis of MSA is closely related to α-synuclein. Therefore, we take k_2 equal to 0.75 and the model is robust to the selection of k_2 values.

$$\Delta A_i(t) = k_1\left(1 - e^{-r_i(t)\Delta t}\right) + k_2 \sum_j \frac{w_{ij}}{\sum_j w_{ij}} (1 - e^{-r_i(t-1)\Delta t}) \tag{9}$$

III. RESULTS

In Fig.1, independent component analysis of DBM is used to replicate the distribution of atrophy in Parkinson's disease (PD). The 17th component was consistent with the PD-ICA network proposed by Zeighami et al. [15]. The result further supports the network transmission mechanism of PD and providing strong evidence for misfolding and the transmission of neurotoxic proteins.

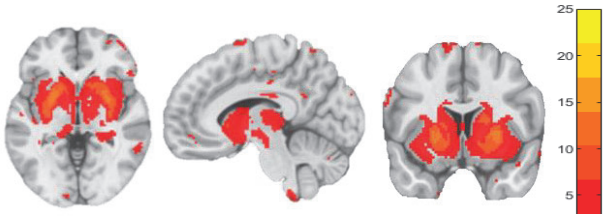

Figure 1 PD-ICA network

(Selected slices in MNI space z = −2, x = −8, y = 10, p>3)

The model has two fixed points, as shown in Fig. 2. The blue line means that M decreases with N and the green line means that N increases with M. The red dots represent two fixed points and the vector field (arrows) denotes the direction of the gradient at each position (i.e., the system is going to move in the direction of the arrow). The final positions of which will not be affected by the initial conditions of (N_i, M_i), including the choice of seed region and seeded misfolded agents.

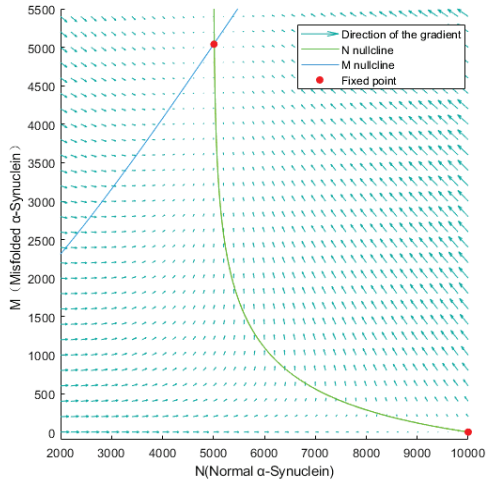

Figure 2 An illustration of the phase plane

A definition of disease epicenter is the seed node most likely to initiate a disease outbreak. We posited that the probability of triggering an outbreak indicates the plausibility of acting as an epicenter. Therefore, we quantified the spread threshold for each seed region, i.e., the minimally required injection amount of misfolded α-synuclein to initiate an outbreak. The spread threshold of substantia nigra is the lowest (spread threshold was inverted by $-log_{10}$), as shown in Fig.3. Therefore, substantia nigra is the most plausible seed region to initiate an epidemic spread. This is consistent with the notion that substantia nigra acts as the epicenter for propagation to the supratentorial central nervous system [16] and is generally one of the earliest regions to display neuronal loss in clinically overt PD.

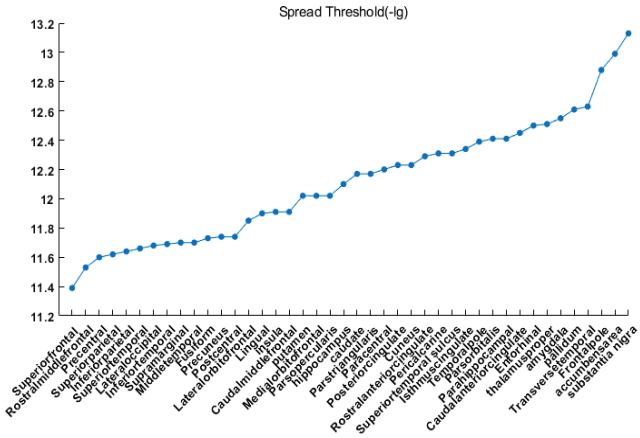

Figure 3 Spread threshold (spread threshold was inverted by $-log_{10}$)

To investigate the relationship between *LRRK2* gene and membrane transport, we use the probability of staying in or entering the brain region to indirectly simulate the process of regulating membrane transport. We measured the fitting degree of the model by Pearson correlation of these two conditions with empirical atrophy respectively. The fitting value of the classic SIR model was 0.60 ($r = 0.60$, $p = 2.21 \times 10^{-5}$,the 95% confidence interval as being between 0.55 and 0.85), which reveals the validity of the classical model. The *LRRK2* gene regulated protein entering cells, and the fitting value of the model remained 0.69 ($r = 0.69, p = 4.05 \times 10^{-7}$). And the *LRRK2* gene regulated protein secretion out of cells, the fitting value of the model could reach 0.73 ($r = 0.73,\ p = 3.33 \times 10^{-8}$, the 95% confidence interval as being between 0.36 and 0.76), as can be seen from Fig.4. Models with *LRRK2* gene expression regulation were fitter than the classic SIR model, which supported the theory of *LRRK2* gene expression modulating cell-to-cell propagation of misfolded proteins. Specifically, we speculated that *LRRK2* gene was more likely to regulate protein secretion because of its higher fitting degree, but there was no evidence at the molecular level to prove its authenticity.

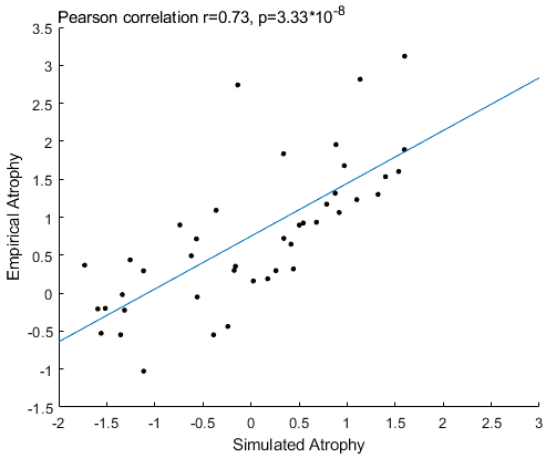

Figure 4 Model fitting in the case of *LRRK2* gene regulating α-synuclein secretion

IV. Discussion

Neuropathologic studies of PD are unable to explain the mechanism of the transformation from local 'protein disease' to large-scale network dysfunction and atrophy yet [16,17]. Meanwhile, the lack of relevant drugs to track α-synuclein makes it impossible to map this protein's propagation directly. Therefore, it is of great important to study the protein transmission model to explore the pathological process of PD. Previous study showed high potential of SIR model describing α-synuclein propagation in explanation of PD pathology [3]. By integrating *LRRK2*, the computational model proposed by current study enhanced the fitting accuracy significantly (r=0.73 compared to 0.6), showing *LRRK2* plays an important role in α-synuclein propagation. Furthermore, the disease transmission in this model is driven by non-cellular autonomic factors (structural connectivity) and cellular autonomic vulnerability (regional gene expression data) supporting the development of PD as a multi-factor process [15,18].

In this study, we quantified the transmission threshold to reveal that the substantia nigra may be the center of transmission of the disease [16], which is consistent with the concept of substantia nigra as the center of transmission to the central nervous system on the screen [15]. The computational model proposed in this study significantly improved the fitting accuracy of the classic model by integrating *LRRK2*, proving that the *LRRK2* gene does participate in the pathogenesis of PD, and provides a new method for studying the pathogenesis of PD. The research had perfected the modeling of PD pathogenesis. And by comparing the fitting accuracy of *LRRK2* gene regulation protein activity (r = 0.73 compared to 0.69), we speculated that *LRRK2* gene was more likely to be involved in protein secretion, which provided new insights into gene regulation protein.

In sum, our results imply *LRRK2* gene expression may affect the PD pathology propagation in central nervous system [6]. Furthermore, our results suggest *LRRK2* gene would be involved in the process of protein secretion instead of entering the cell. Congruent with previous converging evidence in animal and cellular studies, current study highlights the involvement of *LRRK2* in PD pathology and brings new insights to understanding PD mechanism, which provides a new direction for modeling the pathogenesis of PD in the future. Future work could expand to effects of more misfolded proteins and related neurological disorders, such as tau protein propagation in Alzheimer's disease.

Acknowledgment

This study was supported by grants from National Key Research and Development Program of China (2018YFC1312000), the National Natural Science Foundation of China (61801145), and the Basic Research Project of Shenzhen Science and Technology Program (JCYJ20180306171800589).

References

[1] Brundin P, Melki R. Prying into the prion hypothesis for Parkinson's disease. Journal of Neuroscience.2017; 37(41):9808–9818. https://doi.org/10.1523/JNEUROSCI.1788-16.2017 PMID: 29021298

[2] Walsh DM, Selkoe DJ. A critical appraisal of the pathogenic protein spread hypothesis of neurodegeneration.Nature Reviews Neuroscience. 2016; 17(4):251. https://doi.org/10.1038/nrn.2016.13 PMID:26988744

[3] Zheng YQ, Zhang Y, Yau Y, Zeighami Y, Larcher K, et al. (2019) Local vulnerability and global connectivity jointly shape neurodegenerative disease propagation. PLOS Biology 17(11): e3000495. https://doi.org/10.1371/journal.pbio.3000495

[4] Gitler, A., Chesi, A., Geddie, M. et al. α-Synuclein is part of a diverse and highly conserved interaction network that includes PARK9 and manganese toxicity. Nat Genet 41, 308–315 (2009). https://doi.org/10.1038/ng.300

[5] Agrahari Ashish Kumar,George Priya Doss C,Siva R,Magesh R,Zayed Hatem. Molecular insights of the G2019S substitution in LRRK2 kinase domain associated with Parkinson's disease: A molecular dynamics simulation approach[J]. Journal of theoretical biology,2019. https://doi.org/10.1016/j.jtbi.2019.03.003

[6] Henderson, M.X., Cornblath, E.J., Darwich, A. et al. Spread of α-synuclein pathology through the brain connectome is modulated by selective vulnerability and predicted by network analysis. Nat Neurosci 22, 1248–1257 (2019). https://doi.org/10.1038/s41593-019-0457-5

[7] Di Maio R., Hoffman E. K., Rocha E. M., Keeney M. T., Sanders L. H., De Miranda B. R., et al. (2018). LRRK2 activation in idiopathic Parkinson's disease. *Sci. Transl. Med.* 10:eaar5429. 10.1126/scitranslmed.aar5429. PMID:30045977 PMCID: PMC6344941

[8] Alegre-Abarrategui J., Christian H., Lufino M.M., Mutihac R., Venda L.L., Ansorge O., Wade-Martins R. LRRK2 regulates autophagic activity and localizes to specific membrane microdomains in a novel human genomic reporter cellular model. Hum. Mol. Genet. 2009;18(21):4022–4034. http://dx.doi.org/10.1093/ hmg/ddp346. PMID: 19640926.

[9] Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, et al. The parkinson progression marker initiative (PPMI). Progress in neurobiology. 2011; 95(4):629–635. https://doi.org/10.1016/j.pneurobio.2011.09.005 PMID: 21930184

[10] Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, et al. The WU-Minn human connectome project: an overview. Neuroimage. 2013; 80:62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041 PMID: 23684880

[11] Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012; 489(7416):391. https://doi.org/10.1038/nature11405 PMID: 22996553

[12] Desikan RS, Se´gonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labelingsystem for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest.Neuroimage. 2006; 31(3):968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021 PMID:16530430

[13] Cammoun L, Gigandet X, Meskaldji D, Thiran JP, Sporns O, Do KQ, et al. Mapping the human connectome at multiple scales with diffusion spectrum MRI. Journal of neuroscience methods. 2012; 203 (2):386–397. https://doi.org/10.1016/j.jneumeth.2011.09.031 PMID: 22001222

[14] Keuken MC, Bazin PL, Crown L, Hootsmans J, Laufer A, Mu¨ ller-Axt C, et al. Quantifying inter-individual anatomical variability in the subcortex using 7T structural MRI. NeuroImage. 2014; 94:40–46. https://doi.org/10.1016/j.neuroimage.2014.03.032 PMID: 24650599

[15] Zeighami Y, Ulla M, Iturria-Medina Y, Dadar M, Zhang Y, Larcher KMH, et al. Network structure of brain atrophy in de novo Parkinson's disease. Elife. 2015;4.

[16] Kordower JH, Chu Y, Hauser RA, Freeman TB, Olanow CW. Lewy body‐like pathology in long-termembryonic nigral transplants in Parkinson's disease. Nature medicine. 2008; 14(5):504. https://doi.org/10.1038/nm1747 PMID: 18391962

[17] De Pablo-Fernández E, Lees AJ, Holton JL, Warner TT. Prognosis and Neuropathologic Correlation of Clinical Subtypes of Parkinson Disease. JAMA Neurol. 2019 Apr 1;76(4):470-479. doi: 10.1001/jamaneurol.2018.4377. PMID: 30640364; PMCID: PMC6459129.

[18] Surmeier DJ, Obeso JA, Halliday GM. Selective neuronal vulnerability in Parkinson disease. Nature Reviews Neuroscience. 2017; 18(2):101. https://doi.org/10.1038/nrn.2016.178 PMID: 28104909